Skin resident T cells

dos Reis Matos, T.

Citation for published version (APA):
Mass cytometry analysis tools for decrypting the complexity of biological systems

Matos TR, Liu H, Ritz J

Journal of Investigative Dermatology 2017 May;137(5):e43-e51
ABSTRACT

Mass cytometry by time-of-flight (CyTOF) experiments allow analysis of over 40 functional and phenotypic cellular markers simultaneously at the single-cell level. The data dimensionality escalation accentuates limitations inherent to manual analysis, as being subjective, labor-intensive, slow, and often incapable of revealing the detailed features of each unique cell within populations. The subsequent challenge of examining, visualizing, and presenting mass cytometry data has motivated continuous development of dimensionality reduction methods. As a result, an increasing recognition of inherent diversity and complexity of cellular networks is emerging, with the discovery of unexpected cell subpopulations, hierarchies, and developmental pathways, such as those existing within the immune system. Here, we briefly review some frequently used and accessible mass cytometry data analysis tools, including Principal Component Analysis (PCA), Spanning-tree Progression Analysis of Density-normalized Events (SPADE), t-distributed stochastic neighbor embedding (t-SNE)-based visualization (viSNE) and Automatic Classification of Cellular Expression by Non-linear Stochastic Embedding (ACCENSE), and Cluster Identification, Characterization, and Regression (CITRUS). Mass cytometry together with these innovative analytic tools have the power to lead to key discoveries in investigative dermatology, including but not limited to identifying signaling phenotypes with predictive value for early diagnosis, prognosis, or relapse, and a thorough characterization of intratumor heterogeneity and disease-resistant cell populations, that may ultimately unveil novel therapeutic approaches.

SUMMARY POINTS:

- New methods are being continuously developed to analyze and best represent multi-dimensional, complex CyTOF data.
- Principal Component Analysis (PCA) provides a visualization of the data in 3D space and identifies the parameters with the most variance among the dataset.
- Spanning-tree Progression Analysis of Density-normalized Events (SPADE) clusters cells into a minimum-spanning hierarchical tree for 2D visualization.
- In t-distributed stochastic neighbor embedding (t-SNE)-based visualization (viSNE) and Automatic Classification of Cellular Expression by Non-linear Stochastic Embedding (ACCENSE), each single cell data point has a unique location in a 2D representation, reflecting the cells’ immunophenotypic similarity or differences in high dimensional space.
- Cluster Identification, Characterization, and Regression (CITRUS) identifies cellular features that correlate to an experimental endpoint of interest.
INTRODUCTION

New methods are being developed to examine, visualize, and present the multi-dimensional complexity of cellular function, identity, and the role of individual cells within biological systems. Mass cytometry using time-of-flight mass spectrometry (CyTOF) currently has the capacity to allow investigation of 40 or more distinct parameters at the single-cell level (Fig. 1). While the technique has not yet been widely adopted within the field of investigative dermatology, it has potential to, for example, allow identification of cell signals for early diagnosis in cutaneous T cell lymphoma, allow early detection or predict relapse in psoriasis and atopic dermatitis, and allow thorough characterization of drug-resistant cell populations in skin cancer, eventually unveiling new therapies.

The large amount of data generated and potential of the technique to delineate rare cell subsets has driven the need to develop dimensionality reduction methods and analysis algorithms to best analyze and represent mass cytometry data. A significant limitation of traditional data clustering methods through biaxial plots and histograms, such as has been used to represent traditional flow cytometry data, is that pre-existing knowledge of the defining markers of each population is required. This limits the ability of researchers to discover unexpected cellular subsets and does not allow examination of system-level phenotypic diversity. Furthermore, manual analysis of individual markers and combinations of markers is a subjective, slow, and labor-intensive process, which results in a significant scalability restriction and can introduce several inherent biases. While CyTOF technology and experimental methodology have been described in detail in previous review\textsuperscript{1,2}, comprehensive understanding is also required with respect to the tools available for analysis of high-dimensional datasets in order to make meaningful use of the results. In this short review, we focus on some of the most commonly used and accessible novel CyTOF data analysis tools, including Principal Component Analysis (PCA), Spanning-tree Progression Analysis of Density-normalized Events (SPADE), t-distributed stochastic neighbor embedding (t-SNE)-based visualization (viSNE) and Automatic Classification of Cellular Expression by Non-linear Stochastic Embedding (ACCENSE), and Cluster Identification, Characterization, and Regression (CITRUS).
The experimental procedure can be separated into multiple steps, including (i) sample preparation, (ii) cell staining, (iii) cell barcoding, (iv) acquisition of CyTOF data, (v) data processing (randomization, normalization and cell debarcoding), and (vi) high-dimensional data analysis. Briefly, cells are isolated from blood or solid tissue samples. After the optional enrichment step, the cell suspension is stained with Cisplatin or rhodium to distinguish dead cells. Principally, cells were probed with surface, intracellular, or intranuclear markers after tetramer staining (if applicable). Cells can be further barcoded by Mass-tag cell barcoding (after cell fixation), or CD45-Pd cell barcoding (before cell fixation) system. Fixed cells are stained with iridium or rhodium DNA-interchelator and resuspended in deionized water for subsequent acquisition on CyTOF. Collected data are converted into FCS file and metal signals are normalized. Debarcoded samples are loaded onto a bioinformative platform (usually in a MATLAB or R environment, or online server such as Cytobank.org) of choice for high-dimensional cytometric analysis. Adapted from Cheng et al., 2016, with permission from Elsevier.

PCA is a well-established and widely used tool to visualize multidimensional data that was adopted to analyze large mass cytometry data sets. PCA identifies those parameters among a certain dataset that present the most variance by generating linear combinations from a large list of parameters into new compound variables (principal components). As a result, the quotient of the relative variation of each principle component over the total variance gives an idea of the effectiveness of each component in separating out data points. In addition, PCA results in models that can be used to project new data points in linear time. For example, it allows graphical visualization of the expression intensity of several functional markers (Y axis) throughout the cell differentiation process (x axis) (Fig. 2). PCA also allows visualization of the data in 3D space, often prominently displaying the first 3 data components of maximal variance. However, this feature can also be a limitation, as it may mask noteworthy biological differences that are more subtle variances in the data. Another constraint is the inherent assumption that the given data are parametric. PCA also represents the data through linear projections, which may not be representative of the inherent structure of the original data. To overcome this constraint, non-linear methods such as t-SNE (described in following sections) were developed for high-dimensional data analysis.
Figure 1. Mass cytometry experiment workflow. The experimental procedure can be separated into multiple steps, including (i) sample preparation, (ii) cell staining, (iii) cell barcoding, (iv) acquisition of CyTOF data, (v) data processing (randomization, normalization and cell de-barcoding), and (vi) high-dimensional data analysis. Briefly, cells are isolated from blood or solid tissue samples. After the optional enrichment step, the cell suspension is stained with Cisplatin or rhodium to distinguish dead cells. Principally, cells were probed with surface, intracellular, or intranuclear markers after tetramer staining (if applicable). Cells can be further barcoded by Mass-tag cell barcoding (after cell fixation), or CD45-Pd cell barcoding (before cell fixation) system. Fixed cells are stained with iridium or rhodium DNA-interchelator and resuspended in deionized water for subsequent acquisition on CyTOF. Collected data are converted into FCS file and metal signals are normalized. De-barcoded samples are loaded onto a bioinformative platform (usually in a MATLAB or R environment, or online server such as Cytobank.org) of choice for high-dimensional cytometric analysis. Adapted from Cheng et al., 2016, with permission from Elsevier.

DIMENSION REDUCTION AND VISUALISATION ALGORITHMS

Principal component analysis
PCA is a well-established and widely used tool to visualize multidimensional data that was adopted to analyze large mass cytometry data sets. PCA identifies those parameters among a certain dataset that present the most variance by generating linear combinations from a large list of parameters into new compound variables (principal components). As a result, the quotient of the relative variation of each principle component over the total variance gives an idea of the effectiveness of each component in separating out data points. In addition, PCA results in models that can be used to project new data points in linear time. For example, it allows graphical visualization of the expression intensity of several functional markers (Y axis) throughout the cell differentiation process (x axis) (Fig. 2). PCA also allows visualization of the data in 3D space, often prominently displaying the first 3 data components of maximal variance. However, this feature can also be a limitation, as it may mask noteworthy biological differences that are more subtle variances in the data. Another constraint is the inherent assumption that the given data are parametric. PCA also represents the data through linear projections, which may not be representative of the inherent structure of the original data. To overcome this constraint, non-linear methods such as t-SNE (described in following sections) were developed for high-dimensional data analysis.
Figure 2. 3D-Principal Component Analysis (PCA) elucidates the phenotypic and functional complexity of CD8+ T cell memory differentiation. In this example, PCA was used to represent simultaneously 25 markers from a single cell sample, hence quantifying the expression of functional markers among several memory CD8+ T cell subsets. (A) The cytometry data set is plotted on the first 3 principal component axes and shown from 3 different perspectives (rotated around the PC2-axis). Cells are gated according to their surface memory markers, naive (green), central memory (Tcm, yellow), effector memory (Tem, blue), and short-lived effector (Tslec, red), which reveal main phenotypic clusters. (B and C) To analyze only memory cells, cells were gated to exclude the naive compartment (cells with low value for PC1). The average expressions for each phenotypic (B) and functional (C) parameter were normalized and plotted as a function of normalized PC2 values. In this way, the phenotypic progression of CD8+ memory T cells are represented by the x-axis (0 – early differentiated Tcm, 1 – mature Tslec) and the y-axis represents the average expression of each marker. The functional progression of these numerous markers during CD8+ T cell memory differentiation would not be possible by conventional manual gating. Reprinted from Newell et al., 2012, with permission from Elsevier.

Newell at al. (2012) used PCA to represent simultaneously 25 markers from a single cell sample, hence quantifying the expression of functional markers among several memory CD8+ T cell subsets. This representation method displayed a greater phenotypic and functional complexity among CD8+ T cells than previously appreciated (Fig. 2). The holistic study of many functional and phenotypic markers and their expression levels through several differentiation subsets would not be possible by conventional manual analysis. This study also observed that subsets that develop in response to different viruses have distinct combinatorial patterns of cytokine expression, showing the remarkable flexibility of CD8+ T cells in responding to pathogens. A similar approach could be used to study the diversity and complexity of skin-specific T cells or innate lymphoid cells (ILCs), complementing the recent in situ topographic characterization of ILCs in human skin.
Spanning-tree Progression Analysis of Density-normalized Events

In contrast to PCA that draws out the underlying variance within a dataset, the goal of clustering algorithms is to visualize common patterns within data sets. In the context of immune-phenotyping, automatic clustering algorithms aim to define the most prevalent cell populations by clustering cells based on markers expression similarity.\textsuperscript{7}

SPADE was the first algorithm specifically developed for mass cytometry data analysis that includes both clustering and dimension reduction. In SPADE, cells are clustered into a hierarchical tree shape for 2D visualization.\textsuperscript{4,8} Each cluster of a SPADE tree is graphically represented by a circular node, in which the node size symbolizes the frequency of data points within that population (number of hits) and the node color shows the signal intensity of a selected marker (intensity of hit). Each node is connected to related nodes by a minimum-spanning tree (MST) algorithm. MST is a classic computer algorithm that searches for a tree-like graph through a set of spatial nodes by using the least possible number of connections. Such a tree might mimic a relational map of immune cell types in their cell development. This method allows comparison between expression of markers among clusters and across distinct samples in a dimension-reduced big-picture view of diverse cell populations (Fig. 3). A limitation of SPADE is that it cannot reproduce the same representation of results when the same data set is analyzed more than once, since the algorithm involves several stochastic steps. The rigid connections established within the data representation structure (graph) may mislead the positioning of some cell nodes, possibly obscuring the underlying biology. Lee et al. (2015) recently utilized CyTOF and SPADE to define in detail the phenotype and functional characteristics of distinct subsets of nasal dendritic cells in mice.\textsuperscript{9} SPADE clearly portrayed a map of the vast heterogeneity of nasal dendritic cells and identified new subsets that were not perceivable by manual gating on the basis of canonical marker expression. Moreover, it enabled simultaneous comparison of all subsets prior to and after stimulating factor (FMS)-related tyrosine 3 kinase ligand treatment, revealing which subsets became functionally active and/or expanded in number of cells.
Figure 3. Flowchart of a SPADE tree construction and result visualization. This example illustrates how a SPADE tree is generated from raw cytometry data and how it displays the data set results. (i) The cytometry dataset analysis by 2-parameters detects one rare population and three abundant populations. (ii) The original data is subjected to density-density down sampling. (iii) Agglomerative clustering results of the down sampled cells. (iv) Minimum spanning tree algorithm connects the cell clusters. (v) Colored SPADE trees. Nodes are colored by the median intensities of markers, allowing visualization of the markers expression across the numerous heterogeneous cell populations. The final SPADE tree representation enables to determine how many different subsets are present within a dataset, the relative population density (size), the expression of various markers (color) within each subset, and the relationship among subsets (links). Reprinted by permission from Macmillan Publishers Ltd: Nature biotechnology (Qui et al., 2011.), copyright 2011.
t-distributed stochastic neighbor embedding-based visualization and Automatic Classification of Cellular Expression by Non-linear Stochastic Embedding

t-SNE allows visualization of high-dimensional data through a non-linear reduction where each data-point is given a location in a 2D or 3D map. Newer t-SNE-based strategies have been developed to visualize complex mass cytometry data: viSNE and ACCENSE. In contrast to PCA, these algorithms effectively capture nonlinear relationships among the data and, unlike SPADE, they do not cluster cells into exclusive nodes. Each single cell data point has a unique location in a 2D representation, similar to a biaxial scatter plot, reflecting their proximity in high dimensional space. viSNE is a graphical user interfaced tool based on t-SNE, while ACCENSE has extended t-SNE with clustering of the resulting 2D scatter data into density-based clusters. Thus, close proximity between any two cells is based on their immunophenotypic similarity pre-defined by input markers. For example, to generate a map of the various memory T-cell subpopulations markers such as CD31, CD45RA, CD45RO, CCR7, or CD62-L can be selected. The algorithm then clusters cells according to their similarity within these markers.

A color scale can then be used to visualize each marker’s relative expression in the population. However, this type of representation may obscure subtle density differences among cell populations. Another drawback of these algorithms is the inability to currently analyze large numbers of different samples simultaneously. For example, the current version of the commercial software Cytobank can analyze up to 2 million cells.

Healthy and cancerous bone marrow samples were recently studied using viSNE. Healthy samples were graphically represented and contrasted with cancerous samples that exhibited an abnormal map. Marker expression patterns were analyzed from diagnosis to relapse, allowing identification of disease-specific subsets (Fig. 4).

ACCENSE helped to further characterize CD8+ T cell subpopulations and identify new subsets that were not noticeable on biaxial plots (Fig. 5).
Figure 4. viSNE creates a map of the immune system. (A) In this example, viSNE projects a one-dimensional curve embedded in 3D (left) to 2D (right). The color gradient demonstrates that nearby points in 3D remain close in 2D. (B) Application of viSNE to a healthy human bone marrow sample. viSNE automatically separates cells based on their subtype. Each point in the viSNE map represents an individual cell and its color represents its immune subtype based on independent manual gating. The axes are in arbitrary units. (C) Biaxial plots representing the same data shown in B. Select subpopulations are shown with canonical markers where the square color matches the subtype in B. The actual gating used is more complex and uses a series of biaxial plots for each population. Note, unlike B, these plots do not separate between all subtypes in a single viewpoint. (D) The same viSNE map represented in B, but this time each cell is colored based on CD11b expression. Gated cells are all CD33 high and show a CD11b (maturity) gradient. Many of these cells were not classified as monocytes by manual gating (grey cells B). Reprinted by permission from Macmillan Publishers Ltd: Nature biotechnology (Amir et al., 2013), copyright 2013.
Figure 5. ACCENSE quantifies distinct subsets of CD8+ T cells. In this example, ACCENSE was able to stratify cells into phenotypic subsets by the local probability density of cells (ACCENSE map location of hits) based on the 35 markers studied. It identified a novel subset with a unique multivariate phenotype, which is not distinguishable on a biaxial plot of markers. (A) Illustration of a sample mass cytometry dataset. Rows correspond to different cells whereas columns correspond to the different markers. Entries correspond to transformed values of mass–charge ratios that indicate expression levels of each marker. (B) Biaxial plots exemplify the manual gating approach to identify cell subsets. (C) The 2D t-SNE map of CD8+ T cells, where each point represents a cell (M = 18,304) derived by down-sampling the original dataset. (D) A composite map depicting the local probability density of cells as embedded in panel C, computed using a kernel-based transform. Local maxima in this 2D density map represent centers of phenotypic subpopulations and were identified using a standard peak-detection algorithm. viSNE also generates a similar 2D t-SNE map, however each subpopulation has to be identified and demarcated manually, while ACCENSE automatically defines subsets based on local cell density in the map. Reprinted from Shekhar et al., 2014, in accordance with permission rights to reprint material published in PNAS.

CLUSTERING AND CORRELATION WITH CLINICAL OUTCOME

Cluster Identification, Characterization and Regression

CITRUS aims to not only showcase biological mechanisms and hierarchies and define population subsets, similar to previously described algorithms, but also to identify cellular features that correlate to an experimental endpoint of interest. Bruggner et al. (2014) showed that CITRUS accurately identified subsets of cells associated with acquired immune deficiency syndrome-free survival in human immunodeficiency virus-infected patients.13 CITRUS clusters cells based on similarity of cell markers. Such methods can delineate clusters of cells with specific behavioral characteristics such as those that have activated specific signaling pathways, which can potentially be associated with clinical outcomes such as time to recovery or overall survival after a drug treatment or surgery. By specifying the outcome of interest for each sample from the vast cytometry data uploaded, regularized, statistic algorithms can recognize cells exhibiting behavior
predictive of outcome. The phenotype and behaviors of each cluster are delineated through distinct data representations including conventional biaxial plots that can result in a predictive model for the analyses or validation of future samples. In order to run correlation analysis with enough statistical power, CITRUS requires input from more than 8 samples for each group. Gaudillière et al. (2014) applied this method to monitor patients undergoing hip replacement surgery to characterize the phenotypical and functional immune responses predictive of recovery from surgical trauma (Fig. 6). Since the patient clinical histories of post-surgical recovery were known, CITRUS identified intracellular signaling markers that strongly correlated with recovery from fatigue, functional hip impairment, and pain after surgery. Likewise, CITRUS could be used to analyze immune cell populations and molecular markers from patients with inflammatory or autoimmune diseases before and after receiving biologic treatments. It could reveal which cell subset or markers are predictive of therapeutic outcome, allowing early adjustments to treatment protocols in patients predicted to be non-responders, reducing side effects and costs.

COMMERCIAL SOFTWARE AND PUBLIC PROGRAMS

Following mass cytometry data acquisition, each respective sample must be deconvoluted if samples were initially barcoded and normalized based on calibration beads. CyTOF data can be exported in the form of FCS files, which can be analyzed by standard flow cytometry software such as FCS Express and FlowJo, or by using cloud-based analysis tools, such as Cytobank (www.cytobank.org). Cytobank allows transfer and storage of multiple CyTOF FCS files, attachment of related data including protocols, presentations, annotations and images, and the sharing of data and analysis with chosen collaborators. Additional analytical tools may be needed including but not limited to: dose response, heat maps, SPADE, viSNE, CITRUS, and dot and histogram overlays. viSNE can be publicly licensed to academic users as a matlab-based tool from the Dana Pe’er Lab of computational Systems of Biology webpage (http://www.c2b2.columbia.edu/danapeerlab/html/software.html). The Nolan Laboratory also publically offers some of these algorithms, such as SPADE and CITRUS (https://github.com/nolanlab).

ACCENSE is freely offered at (http://www.cellaccense.com) and a PCA algorithm is included in the open-sourced R basic package. These analysis tools may provide different types of biological information about the same data set, making it acceptable to concurrently use distinct tools in a single experiment (Table 1). Nevertheless, the accessibility to CyTOF analysis tools continues to improve, with numerous groups developing similar and new analysis methods for CyTOF data.
Mass cytometry analysis tools for decrypting the complexity of biological systems

Predictive of outcome. The phenotype and behaviors of each cluster are delineated through distinct data representations including conventional biaxial plots that can result in a predictive model for the analyses or validation of future samples. In order to run correlation analysis with enough statistical power, CITRUS requires input from more than 8 samples for each group. Gaudillière et al. (2014) applied this method to monitor patients undergoing hip replacement surgery to characterize the phenotypical and functional immune responses predictive of recovery from surgical trauma (Fig. 6). Since the patient clinical histories of post-surgical recovery were known, CITRUS identified intracellular signaling markers that strongly correlated with recovery from fatigue, functional hip impairment, and pain after surgery. Likewise, CITRUS could be used to analyze immune cell populations and molecular markers from patients with inflammatory or autoimmune diseases before and after receiving biological treatments. It could reveal which cell subset or markers are predictive of therapeutic outcome, allowing early adjustments to treatment protocols in patients predicted to be non-responders, reducing side effects and costs.

Following mass cytometry data acquisition, each respective sample must be deconvoluted if samples were initially barcoded and normalized based on calibration beads. CyTOF data can be exported in the form of FCS files, which can be analyzed by standard flow cytometry software such as FCS Express and FlowJo, or by using cloud-based analysis tools, such as Cytobank (www.cytobank.org). Cytobank allows transfer and storage of multiple CyTOF FCS files, attachment of related data including protocols, presentations, annotations and images, and the sharing of data and analysis with chosen collaborators. Additional analytical tools may be needed including but not limited to: dose response, heat maps, SPADE, viSNE, CITRUS, and dot and histogram overlays. viSNE can be publicly licensed to academic users as a matlab-based tool from the Dana Pe'er Lab of computational Systems of Biology webpage (http://www.c2b2.columbia.edu/danapeerlab/html/software.html). The Nolan Laboratory also publically offers some of these algorithms, such as SPADE and CITRUS (https://github.com/nolanlab). ACCENSE is freely offered at (http://www.cellaccense.com) and a PCA algorithm is included in the open-source R basic package. These analysis tools may provide different types of biological information about the same data set, making it acceptable to concurrently use distinct tools in a single experiment (Table 1). Nevertheless, the accessibility to CyTOF analysis tools continues to improve, with numerous groups developing similar and new analysis methods for CyTOF data.

Figure 6. Overview of Citrus. Citrus enables correlation of the multitude of cellular parameters studied with clinical outcomes information. Cells from all samples (i) are combined and clustered using hierarchical clustering (ii). Descriptive features of identified cell subsets are calculated on a per-sample basis (iii) and used in conjunction with additional experimental metadata (iv) to train a regularized regression model predictive of the experimental endpoint (v). Predictive subset features are plotted as a function of experimental endpoint (vi), along with scatter or density plots of the corresponding informative subset (vii). In this example, the abundance of cells in subset A was found to differ between healthy and diseased samples (vi; subset A abundance in healthy patients and diseased patients). Scatter plots show that cells in subset A have high expression of marker 1 and low expression of marker 2 relative to all other cells (shown in gray). In this study, Citrus identified T-cell subsets whose abundance is predictive of AIDS-free survival risk in patients with HIV. Reprinted from Bruggner et al., 2014, in accordance with permission rights to reprint material published in PNAS.
<table>
<thead>
<tr>
<th>Algorithm name</th>
<th>Type of information</th>
<th>Advantages</th>
<th>Limitations</th>
<th>Reference</th>
</tr>
</thead>
</table>
| PCA            | Parameters with most variance within dataset | - Visualization in 3D space | - May miss subtle variances within data  
- Assumes that data is parametric  
- Data representation is restricted to linear projections | Jackson, 1991; Newell et al. 2012 |
| SPADE          | Cell population hierarchies | - Delineates the presence of rare cell types  
- Can compare clusters and expression markers among cell subsets and across samples | - Lacks reproducibility  
- Rigid structure connectivity | Qui et al. 2011 |
| t-SNE based viSNE and ACCENSE | Cell subset heterogeneity | Single-cell representation (without clustering) | - Unfeasible to analyze vast numbers of cells* | Amir et al. 2013; Shekhar et al. 2014 |
| CITRUS         | Allows correlation between sample and clinical outcome | Correlation to experimental endpoint of interest | - More than 8 samples per group are required | Bruggner et al. 2014 |

Table 1. Analysis algorithms for mass cytometry data. Principal Component Analysis (PCA), Spanning-tree Progression Analysis of Density-normalized Events (SPADE), t-distributed stochastic neighbor embedding (t-SNE)-based visualization (viSNE) and Automatic Classification of Cellular Expression by Non-linear Stochastic Embedding (ACCENSE), and Cluster Identification, Characterization, and Regression (CITRUS). *The cloud-based software Cytobank viSNE can currently handle up to 2 million cells.
CONCLUSIONS

The continuous development and enhancement of analysis tools for mass cytometry expands our ability to study complex and heterogeneous biological systems at the level of individual cells. This enables our understanding of the progression and development of healthy and pathologic cells, such as in psoriasis, atopic dermatitis, and vitiligo. For example, why do some lesions only reoccur in the same anatomical site? Why are some areas of the body more commonly affected? What distinguishes pathological cells in a stable versus a progressive disease? Disease-specific cell subsets can be identified, characterized, monitored during treatment, and perhaps screened for early biomarkers predictive of relapse risk. Differences may be revealed among cells responsible for the clinical heterogeneity of cutaneous T cell lymphoma, ultimately unveiling disease-specific biomarkers and personalized novel therapeutic approaches. Innovative therapies can be studied to specifically target malignant cell populations resistant to conventional treatments, such as in melanoma. Ultimately, the goal of this article was to demystify the developing tools for mass cytometry and its data analysis so these technologies may be adopted, and results understood to address these and other important research questions in the coming years.

ACKNOWLEDGMENTS

We would like to thank Dr. Jodi L. Johnson for helpful comments, critical reading of the manuscript, and editorial assistance.
REFERENCES